(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/80013 A1

(51) International Patent Classification⁷: G06F 12/00, 13/00

(21) International Application Number: PCT/IL01/00309

(22) International Filing Date: 4 April 2001 (04.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/198,064 18 April 2000 (18.04.2000) US

(71) Applicant *(for all designated States except US)*: STORE-AGE NETWORKING TECHNOLOGIES [IL/IL]; Gutwirth Science Center, Technion City, 32000 Haifa (IL).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: NAHUM, Nelson [IL/IL]; Morad Hayasmin 4, 34762 Haifa (IL).

(74) Agent: LOWY, Avi; P.O. Box 6202, 31061 Haifa (IL).

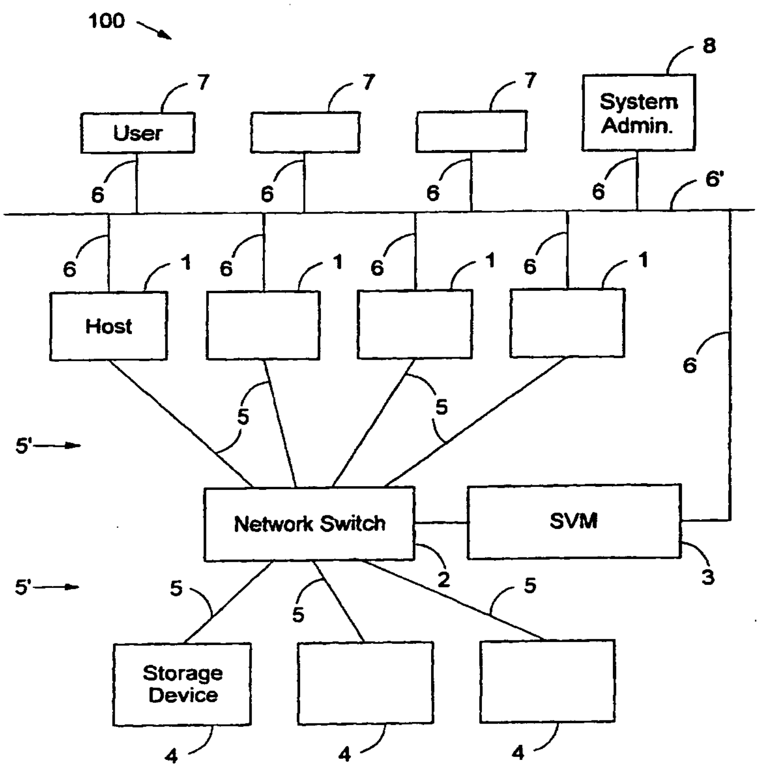
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

[Continued on next page]

(54) Title: STORAGE VIRTUALIZATION IN A STORAGE AREA NETWORK



(57) Abstract: A distributed architecture for the virtualization of storage capacity in a Storage Area Network (SAN) and for the management of virtual and physical memory is described. There is provided a virtualization software computer program consisting of two portions, namely virtualization and translation, each portion residing in a different location but both portions operating interactively. A SAN coupling an array of hosts (1) via a Network Switch (2) to an array of storage devices (4) is equipped with a Storage Virtualization Manager (3). The SVM operating the virtualization computer software handles physical storage capacity virtualization and metadata management. The Network Switch routes storage I/O operations between the hosts and the storage devices, while the translation software resides in a processor, in either a host or elsewhere on the SAN. Although the Network Switch and the SVM decouple tasks to relieve load and prevent bottlenecks, practical implementation permits to design the Network Switch, the processor(s) operating the virtualization program, and the SVM in many configurations spanning from distributed to integrated packaging. The virtualization software also supports real time configuration adaptation of changes occurring in the configuration of the array of hosts and of storage devices of the SAN.

STORAGE VIRTUALIZATION IN A STORAGE AREA NETWORK

Technical field

The present invention relates to the field of Fibre Channel based Storage Area Networks (SAN) in general and more particularly, to storage virtualization and to storage management in a Storage Area Network.

Definitions:

A Host is a computer or a server.

A Storage Area Networks (SAN) consists of an array of host computers and an array of storage devices coupled to a network by a Network Switch.

A Virtual Volume is, or Virtual Volumes are, a list of physical storage areas or Stripe Areas concatenated and presented to a host computer as a single Virtual storage device.

A Stripe Set is a group of physical storage devices treated as one large physical storage device. The capacity of the Stripe Set is defined as the capacity of the smallest physical device multiplied by the number of the devices in the Stripe Set.

Metadata is the data pertaining to the mapping of the Virtual Volumes.

A Fabric is a Fibre Channel Network, or FC Network.

A Storage Pool comprises at least one or more concatenated storage devices.

A Physical Device is a Logical Unit (LU) of a Storage System. Logical units are identified by the 8 bytes of their World Wide Name (WWN) and by 8 bytes of the LU.

Background Art

The present invention relates to the management of physical data storage resources in Storage Area Networks (SAN) consisting of an array of host computers and an array of storage devices coupled in a storage network by a Network Switch. All I/O operations of each host out of the array of hosts of the SAN are processed centrally by the Network Switch for routing to the array of storage devices. In addition, since the Network Switch also centrally manages memory storage allocation, running both tasks of routing I/O operations and managing storage allocation often create bottlenecks preventing timely operation of the SAN.

Storage systems may be related to as pertaining to distinct generations. A first generation of storage system dealt with a single host having a file system that communicated directly with a single storage device. For example, U.S. Patent No. 5,983,316 disclosed by Norwood, divulges a first generation implementation of virtualization for a single host and a plurality of storage devices. U.S. Patent No. 5,404,478 divulged by Arai et al., also describes virtualization and its implementation for a multiplicity of host and storage device pairs, thus again for a one to one relation of the first generation of storage virtualization. The second generation of storage systems consisted of a host with a single operating system able to handle RAIDs (Redundant Array of Independent Disks), which may

It is an additional object of the present invention to provide a method for configuring both the Network Switch and the SVM for incorporation into a joint housing. Accommodating one chosen host of the array of hosts for incorporation therein of the SVM, with the SVM being configured for operative association with the array of hosts and for coupling thereto via the user network, is also possible. Similarly, configuring a host coupled to the user network to operate the configuration computer program in operative association with the Network Switch and with the translation portion of the virtualization computer program operating on at least one processor coupled to the storage network is also practical.

It is a supplementary object of the present invention to provide a method for storage virtualization in a Storage Area Network (SAN) comprising an array of hosts (1) coupled to an array of storage devices (4) having a storage capacity. An object of the present invention also comprises an Enhanced Network Switch (2E) operative for routing storage I/O operations between the array of hosts and the array of storage devices. It is included that the array of hosts, the array of storage devices and the Enhanced Network Switch are being coupled together in a storage network (5'), and the array of hosts are being coupled to a plurality of user workstations (7) in a user network (6'). The method is characterized by comprising the steps of forming an Upgraded Network Switch (UNS) (2*) to upgrade the Enhanced Network Switch (2E). The UNS is integrating an adjusted Storage Virtualization Manager (SVM) (3/500) configured for and operative for virtualization of the storage capacity and for managing metadata, the (UNS) comprising a first Enhanced Network Switch portion and a second adjusted SVM portion. The method further comprises coupling the second adjusted SVM portion by a storage network link (5) to the first Enhanced Network Switch portion and by a user network link (6) to the user network. This is followed by operating the second adjusted SVM portion for virtualization of the storage capacity and for managing metadata, whereby virtualization of the storage capacity and managing of metadata are decoupled from routing storage I/O operations.

It is yet an object of the present invention to provide a method for coupling a plurality of ports of the Enhanced Network Switch comprising coupling a first portion of ports (1*) to the hosts (1) and a second portion of ports (4*) to the storage devices (4), and coupling by storage network link (5) to each port of the first portion of ports and to each port of the second portion of ports, respectively, at least one host and at least one storage device, and coupling a processing instance (66) to each port of at least the first portion of ports, and to the second adjusted SVM portion.

It is still an object of the present invention to provide a method for operating a virtualization computer program comprising a first configuration portion operative on the second adjusted SVM portion and a second translation portion operative on the processing instance coupled to each port of the first portion of ports. Evidently, the method also operating independently the second translation portion of the virtualization computer program on the processing instance coupled to each port of the first portion of ports. Clearly, the method furthermore running

It is still a further object of the present invention to provide a method for operating a virtualization computer program comprising a first configuration portion operated by the adapted SVM and a second translation portion operated by the processing instance coupled to each port of the first portion of ports. Evidently, the method is also for operating the first configuration portion and the second translation portion of the virtualization computer program in interactive operative association.

It is moreover a further object of the present invention to provide a method for operating the first portion of the virtualization computer program to support real time configuration adaptation of the at least one SAN, in response to a configuration change occurring in the array of hosts and/or in the array of storage devices.

It is likewise a further object of the present invention to provide for a method for operating computer program control functions comprised in the virtualization computer program for management of storage virtualization and for configuration management of both the array of hosts and the array of storage devices of the at least one SAN. Possible is also enabling a System Administrator to manage the computer program control functions by operating a workstation coupled to the user network. Moreover, the method allows for managing the computer program control functions in operative association with at least one user and/or storage application computer program operating on a host of the array of hosts of the at least one SAN.

It is nevertheless a further object of the present invention to provide for a system and for a storage virtualizer for storage virtualization in a Storage Area Network (SAN) comprising an array of hosts (1), an array of storage devices (4) with a storage capacity. The Network Switch (2) is operative for routing storage I/O operations between the array of hosts and the array of storage devices. The array of hosts is being coupled to the array of storage devices via the Network Switch in a storage network (5'), and the array of hosts is being coupled to a user network (6), comprising a plurality of user workstations (7). The system and the a storage virtualizer are characterized by comprising a Storage Virtualization Manager (SVM) (3) coupled by a storage network link (5) to the Network Switch on the storage network. The SVM is coupled by a user network link (6) to the user network, and the SVM is configured for and operating for virtualization of the storage capacity and for managing metadata, whereby virtualization of the storage capacity and managing metadata are decoupled from routing storage I/O operations.

It is another object of the present invention to provide for a system and for a storage virtualizer for a virtualization computer program comprising a first configuration portion running on the SVM and a second translation portion operating in association with at least one host of the array of hosts. Each host of the array of hosts is operating the second translation portion of the virtualization software computer program, and/or there is a processor associated with at least one host of the array of hosts for operating the second translation portion of the virtualization computer program.

comprising a first configuration portion operative on the second adjusted SVM portion and a second translation portion operative on the at least one processor coupled to each port of the first portion of ports. This further comprises each at least one processor being coupled to each port of the first portion of ports independently operating the second translation portion of the virtualization computer program. In complement, there is a real time configuration adaptation of the SAN in response to a configuration change occurring in the array of hosts and/or the array of the storage devices, the real time configuration adaptation being supported by operation of the virtualization computer program.

It is nevertheless another object of the present invention to provide for a system and for a storage virtualizer for computer program control functions comprised in the virtualization computer program for management of storage virtualization and for configuration management of both the array of hosts and the array of storage devices. This includes a System Administrator for managing the computer program control functions by operating a workstation coupled to the user network and at least one user and/or storage application computer program for managing the computer program control functions.

It is an additional object of the present invention to provide for a system and for a storage virtualizer for storage virtualization of at least one Storage Area Network (SAN) comprising an array of hosts (1), an array of storage devices (4) having a storage capacity, and a Enhanced Network Switch (2E). The Enhanced Network Switch is operative for routing storage I/O operations between the array of hosts and the array of storage devices, the array of hosts being coupled to the array of storage devices via the Enhanced Network Switch, with the array of hosts and the array of storage devices and the Enhanced Network Switch being coupled together in a storage network (5'), and the array of hosts being coupled to a user network (6') comprising a plurality of user workstations (7) and further linked to a remote host (1R) via an Internet (80). The system and the storage virtualizer are being characterized by comprising an adapted SVM (3/600) coupled in operative association with the remote host, the adapted SVM being configured for virtualization of the storage capacity and for managing metadata of the at least one SAN via the Internet and the user network. The adapted SVM is being operated for virtualization of the storage capacity and for managing metadata of the at least one SAN via the Internet and the user network, and a coupling for linking the Enhanced Network Switch to the user network, whereby virtualization of the storage capacity and managing of metadata are decoupled from routing storage I/O operations.

It is yet an additional object of the present invention to provide for a system and for a storage virtualizer wherein the Enhanced Network Switch (2E) comprises a plurality of ports having a first portion of ports (1*) for coupling to the hosts (1) and a second portion of ports (4*) for coupling to the storage devices (4), and coupling by storage network links (5) to each port of the first portion of ports and to each port of the second portion of ports, respectively, at least one host of the array of hosts and at least one storage device of the array of storage devices. The system and the storage virtualizer are being characterized by further

Fig. 12 exhibits a flow-chart illustrating the process of changing the permission of a Virtual Volume,

Fig, 13 explains, in flow-chart form, how device additions to the SAN configuration are tracked,

Fig. 14 visualizes how device deletions to the SAN configuration are detected,

Fig. 15 illustrates a second embodiment 200 of the system shown in Fig. 1,

Fig. 16 represents a third embodiment 300, in addition to the embodiments of Figs. 1 and 2,

Fig. 17 divulges a fourth embodiment 400,

Fig. 18 is a diagrammatic representation of a preferred embodiment 500,

Fig. 19 shows a detail of Fig. 18, and

Fig. 20 depicts a last preferred embodiment 600.

Disclosure of the Invention

The present invention, proposes a new distributed architecture for the virtualization of storage in a Storage Area Network (SAN), and for the centralized and uniform management of Virtual Volumes of storage. The method allows a System Administrator to exercise real time uniform management of the Virtual Volumes by software control, without requiring expensive hardware or sophisticated software for implementation. The method is used to create Virtual Volumes of storage representing the available physical storage facilities. One or more Virtual Volumes of memory are then presented to a host and appear to the host as being connected to a storage area. A host thus sends a storage request (either write or read) to a virtual address in a Virtual Volumes(s). The virtual address is then translated into physical storage location(s) where data are saved.

Accordingly, there is provided a virtualization software computer program consisting of two different portions, each portion residing in a different location but both portions remaining in communication with each other.

In broad terms, the first portion of the virtualization software, called the configuration software, is used to create Virtual Volumes of storage and to handle the metadata, i.e., to map and maintain configuration tables of both the virtual and the physical storage. The second portion of the virtualization software, referred to as the translation software, mainly consists of a translator, for the translation of virtual memory addresses into physical memory locations. The virtualization computer program thus virtualizes the storage capacity and manages related metadata.

The virtualization software operates on a system involving a SAN with a Network Switch and a Storage Virtualization Manager (SVM). By definition, a SAN includes an array of host computers coupled via a Network Switch to an array of storage devices, either by Fibre Channel (FC), or by TCP/IP, or by other links, to one or more storage devices. The SAN is linked to a user network, preferably an Ethernet, such as a LAN, a WAN, or the Internet, to which users are connected. A System Administrator, operating a workstation coupled to the user storage area in the form of Virtual Volumes and to map the configuration relating the virtual memory addresses to the physical locations in the array of physical storage devices 4. Each host 1 may derive from the SVM 3 a host-specific table of translation, for translation from virtual storage address into physical storage location. Virtual Volumes created by the SVM 3 are thus presented to the hosts 1, whereas the translation from virtual addresses into physical locations for data storage is processed by the translation software operating in each host in the embodiment 100.

Furthermore, the SVM 3 runs a device-polling mechanism for a continuous status-update, and availability update of the host computers 1 and of the physical storage devices 4 operating on the SAN. The SVM 3 maps the presence, availability and permission level granted to each host computer into a detailed configuration table of allocations also including data about the storage capacity available in the physical storage devices 4.

The SVM 3 also receives configuration-polling requests from an SVM Driver, residing in each host 1, but not shown in Fig. 1. The SVM Driver operates a configuration polling mechanism, launched every few seconds (say 5 to 10 for example), for the detection of changes occurring in the configuration of the SAN. The detected changes are then entered in the configuration tables maintained by the SVM 3. Each SVM Driver derives, for the specific host 1 in which it resides and in relation with the particular Virtual Volume(s) dedicated thereto, a host-specific table of translations and of permissions.

Since systems are dynamic, a system reconfiguration occurs upon addition or deletion of hardware devices, e.g. hosts 1 or storage devices 4. Such changes result either from modifications entered by the System Administrator say for addition of capacity or for maintenance purposes of the SAN, or from hardware failures. In consequence, the configuration of the SAN is prone to alterations caused either by a change in the number of hosts 1, or by an adjustment of the available storage devices 4, or by the intervention of the System Administrator accessing the SVM 3 to command a hardware reconfiguration. The virtualization software computer program adapts the SAN in real time in response to configuration changes of the devices, namely hosts 1 and storage devices 4. This means adaptation of the SAN to the new configuration, including the supporting metadata, tables, and other information required for operation.

When a host 1 initiates an I/O storage command, the address translation is based on the latest update of the translation tables derived from the SVM 3 by the SVM Driver running in that host. Practically, this simple address translation is an undemanding operation that does not impose any processing load on the host 1.

The method described above prevents data transfer bottlenecks in the virtualization appliance, i.e. the SVM 3, by avoiding the creation thereof. Such data blockages occur on standard equipment when I/O commands are handled by a central management device that also operates storage management tasks and routing operations from hosts 1 to physical storage devices 4. With the proposed method, in contrast with the conventional systems, the obligation to pass the data components are attached to the FC-IB 11 but are not described since they are standard and do not add to the explanations.

The SVM Software

Referring now to Fig. 4, the principal software modules of the SVM 3 are depicted as an example for a specific FC embodiment. The programs referred to as Pentium programs are run by the Win NT on the Single Board Computer 10 (SBC 10) while those denominated as 960 programs are run by the i960 RN processor 30 on the FC-IB 11.

The Pentium programs include, amongst others, the Windows NT operating system 41 (Windows NT 41), with a standard TCP/IP stack 42, a web-server 43, and an I2O device driver module 44. The I2O module 44 executes the data communication tasks between the Windows NT 41 and the Ixworks Real Time Operating System 45 (Ixworks RTOS 45) operated by the i960 RN processor 30 on the FC-IB 11.

The 960 programs encompass the FC-IB 11, the Ixworks Real Time Operating System 45 (IX RTOS 45) with an I20 Support module 46 and the following software modules: an FC driver 47, a Disk/HBA driver 48, a Setup module 49, a Storage Manager 50 and an HTML builder 51. The I2O support module 46 is in charge of the data communications exchange with the I2O driver module 44 of the Windows NT 41.

The FC driver module 47, programmed according to the Qlogic firmware specifications, handles the FC software interface for communication of the SVM 3 with the Network Switch 2. In other words, the FC driver module 47 enables FC communication between the SVM 3 and the Network Switch 2, and serves for interaction with the hardware devices of the SAN, such as the hosts 1 and the storage devices 4.

The Disk/HBA driver 48 is needed for communication of the SVM 3 with the storage devices 4 and with the hosts 1, or more precisely, with the HBAs of the hosts 1. The Disk/HBA module 48 operates a device-polling mechanism for detection of the presence, of the availability, and of the access permission to the various hardware devices, host(s) 1 and storage devices 4, of the SAN. The Disk/HBA module 48 is thus able to detect the status parameters per Virtual Volume, including the presence and the authorization level of the hosts 1 and of the storage devices 4. Status parameters, for a host 1 and for storage device 4, include Absent, Present, or Failed. The authorization levels of the storage devices 4 include Read Only, Read/Write and Off-Line. In addition, since the number of hosts 1 and the number of storage devices 4 coupled to the SAN may change due to availability or decision of the System Administrator, the Disk/HBA module 48 is responsive to configuration modifications, such as the addition or deletion of hardware units, as will be explained below.

The device-polling mechanism operated by the Disk/HBA module 48 sends an inquiry command, in this example a SCSI inquiry command, every 30 seconds to all the FC devices of the SAN. The information retrieved describes all the devices connected on the SAN, for the update of the database of the Qlogic FC controller

Actually, the standard HBA driver is replaced by a Virtual Volume Driver but may still be referred to as an HBA driver.

Configuration Table

An example of a configuration table created and managed by the SVM 3 is now shown with reference to the Figs. 5, 6 and 7. Both Figs. 6 and 7 are derived from Fig. 5 and present partial views of the configuration table.

The SVM 3 creates one configuration table for each Volume of virtual memory. All of the configuration tables are maintained in the SVM 3 but each SVM Drive, pertaining to a host 1, retrieves one host-specific configuration table for each allowed Virtual Volume. Therefore, each single operational host 1 maintains at least one configuration table from which translation from virtual address to physical location is derived. An excerpt of such a configuration table is shown in Fig. 5.

The configuration table is retrieved by the SVM Driver of a specific host 1, provided that the at least one HBA connected to that specific host 1 has permission thereto. In Fig. 5 the example relates to a first Virtual Volume labeled as "1: VOLUME Vol#1", as by the first line of Fig. 5.

The first paragraph of Fig. 5 holds information about addresses and storage. In the present example, a series of four storage areas is listed, from pRaid[0] to pRaid[3], each storage area being accompanied by the name of the physical device, such as for example, in the first line of the first paragraph, the name is 200000203710d55c0000000000000000 for pRaid[0].

The second line of the first paragraph, namely qwRaidStartLba[0] = 0, indicates the starting location of storage as being sector 0, in the storage device pRaid[0], and the third line of the same paragraph indicates that the end location is sector number 10240000, or dwNumLba[0] = 10240000. Both the Virtual Volume #1 and the storage device pRaid[0] thus use 10240000 sectors of memory in respectively, virtual memory and physical storage.

Another view of Fig. 5 is illustrated in Fig. 6 where the four storage areas of the Volume Vol#1 appear as four consecutive lines. The four storage areas displayed in Fig 6 are located in three different physical devices. In the second column designated as “Physical Device Name”, line two and line four, both carry the same name 200000203700a9e30000000000000000, and are therefore the same storage device.

The first column of Fig. 6 lists the virtual addresses of the Virtual Volume #1. These virtual addresses are derived from Fig. 5, as follows. The first storage area pRaid[0] in the first line of the first paragraph, starts at sector 0 and ends in sector 10240000. The second storage area pRaid[1] in the fourth line of the first paragraph, thus starts at sector 10240000 and since pRaid[1] of Fig. 5 covers a range from 0 to 20480000 sectors, the starting virtual address sector of the third storage area is evidently the sum of both ranges, thus sector 30720000 and so on. These Virtual Volume Addresses are listed in the first column of Fig. 6 against the names of the Physical Devices shown in the second column of the same Fig.

Management of Virtual Volumes

Centralized management of Virtual Volumes via the SVM 3 is described to explain how Virtual Volumes are created, expanded, deleted and reassigned.

The SVM 3 supports three different management interfaces:

1. A GUI (Graphic Utility Interface) allowing the System Administrator to operate a remote Workstation coupled to the users network,

2. A first Application Program Interface (API1) enabling a user application computer program to manage Virtual Volumes without "human intervention", and

3. A second Application Program Interface (API2) permitting a "storage application" computer programs to manage Virtual Volumes without human intervention.

The GUI for the System Administrator

It was explained above that the SVM 3 has an internal Web-server to operate a Pentium SBC with a small I20 based CGI script, to retrieve HTML pages from the HTLM builder software running in the i960 processor. Therefore, the System Administrator is able to operate from a remote site, using the Internet and a standard web browser to access the SVM 3, by connecting the web browser to the IP address of the web server internal to the SVM 3. In operation, hosts 1 and storage devices 4 are discovered automatically by the SVM 3 whiles Stripe Sets, Storage Pools, and Virtual Volumes need to be created by the System Administrator.

To create a Stripe set, the System Administrator is provided with a "Create Stripe Set" menu. It remains for him to enter a name for the Stripe Set and next, to select the Stripe size and the participating storage devices. As a last step, the System Administrator confirms his choice and the new Set is created. Once created, the Stripe Set may be deleted by use of a "Delete Stripe Set" menu.

Similarly, to define a Storage Pool, the System Administrator turns to the "Create Storage Pool" menu, enters a name for the Storage Pool, chooses the devices or Stripe Sets that will be part of the storage pool and confirms his actions. Once confirmed, a new storage pool is produced and the System Administrator may add more storage devices 4 to the newly created pool, either remove storage devices or entirely delete the whole Storage Pools by taking advantage of various dedicated menus.

In the same manner as previously described, the System Administrator operates a "Virtual Volumes menu" to create the Virtual Volumes, giving those Virtual Volumes a name, dedicating them to a storage pool, entering the storage capacity required for the Virtual Volumes, (in Megabytes) and assigning these Virtual Volumes to one or more host(s) 1. Upon confirmation, the one or more Virtual Volumes are created and added to the configuration tables in the SVM 3. Now that a Virtual Volume exists, it may be expanded at any time simply by adding more storage capacity thereto (either from the same Storage Pool or not) or may be reassigned to one or more different host(s) 1, or even deleted at any time.

flow to step A2 to finish the task by one or more round-again loops. Else, step A8 is reached where a response is sent to the operating system of the host 1, which emitted the I/O request, as a notification that the command in question was successfully completed. Flow control now returns to step A1.

With reference to Fig. 10, the process of creation of a new Virtual Volume is explained regarding handling by the System Administrator. To begin with, a request from a user for the creation of a new Virtual Volume is awaited, in step B1. Once a request arrives from a host 1 to the SVM 3, either via the user network of links 5 or via the storage network links 6, the System Administrator is presented with a screen listing the free space still available in the Storage Pools, in step B2. Next, in step B3, the System Administrator is prompted for a Virtual Volume name and for the memory capacity requested, before proceeding to step B4 or else, the process returns to the previous step. At step B4, the System Administrator is presented with a screen listing the levels of permission of the array of hosts 1. To continue to step B6, the System Administrator must select levels of permission for the hosts 1, as requested in step B5, or else, the process returns to the previous step. In step B6, the System Administrator is provided with a confirmation screen, but step B7 checks whether confirmation is granted. Should confirmation be denied, then the flow reverts to step B2 for another round, but if allowed, then in step B8, the new Virtual Volume is added to the database of configuration tables storing the Virtual Volumes data and status parameters, including presence and permission levels, which are safeguarded in the setup module 49 of Fig. 4. The process ends with step B9 that takes care to indicate the existence of the new Virtual Volume(s) to the specific hosts 1 and to return command to the first step B1. When automatic control (without System Administrator) is desired by an API 1 or an API 2, the user must specify all the necessary parameters

A description of the process for the expansion of a volume is given in Fig. 11, where the first step C1 is the wait for a request from an application program to expand a volume. When such a request is received, either the System Administrator is shown, or the API1 is presented, in step C2, with a full listing of all the free space still available in the storage pools and there is prompting, in step C3, to enter the capacity required. A decision by the System Administrator or a response from the

API 1 allows continuing the process in step C4 where a confirmation screen appears for confirmation, or else, the process returns to the previous step. Upon display of the confirmation screen, the System Administrator is prompted to confirm his decision in step C5, to reach the next stage, namely step C6. Not doing so returns him to the second step C2 to start all over again. At the next before last step C6, the database of configuration tables is updated with the new parameters regarding the expansion of the Virtual Volume, and saved in the setup module 49 of Fig. 4. Finally, at the end in step C7, the hosts 1 are notified of the new Virtual Volume status parameters, and control of the process returns to the first step C1.

Should the device found by Step E3 be an HBA, then the HBA is added to the list and the mechanism checks, in step E7, if the HBA belongs to an already listed host 1 or to a new host. For a previously listed host 1, the WWN of the new HBA is merely added in connection with the proper host 1, as per step E8. However, should the new HBA belong to a new host 1, then the new host 1 is listed in the list of hosts 1, by step E9, and the polling mechanism continues to step 10 to return to the starting step E1.

In the same manner, the polling mechanism also deals with device removal from the SAN, as illustrated in Fig, 14. As before, the poling mechanism starts from the idle state in step F1, and searches for any removed device, in step F2. If no device removal is detected, the polling mechanism returns to the initial idle step F1 for a further search loop.

If step F2 detects that a device was removed, then the search proceeds to step F3 to uncover whether the removal consists of either an HBA or a storage device 4. For a storage device 4, the WWN and the LUN numbers are read, in step F4, to determine, by a next step F5, whether the device is listed or not. A listed storage device 4 is simply marked as absent in step F6, and the mechanism passes to step F12 to return to the beginning step F1. On the contrary, an unlisted device is just deleted from the list, in step F7, before returning to the idle state step, via F12 to F1.

Symmetrically, should step F3 find that the removed device is an HBA, then step F8 marks the HBA as absent and checks to which host 1 the HBA belongs. Should this be the only HBA for that host 1, then the host is designated in step F10, as absent. Otherwise, step F11 labels the host as degraded. Either absent or degraded, the polling mechanism reverts to the idle state in step F1 via step F12.

The SVM 3, operating the configuration computer program thus permits addition and removal of devices from the SAN under constant automatic configuration control in real time and in a flexible manner, without disrupting operations.

Additional Embodiments

Further embodiments of the SAN Virtualization, featuring distributed architecture, will now be presented.

The users 7 and the System Administrator workstation 8 coupled to the user network 6' are deleted from the Figs. 15 to 20 for the sake of simplicity. The configuration of the San remains substantially the same and the same numerals are used to denote the same or similar elements in the Figs 15 to 20.

The term SVM 3 is used hereafter to denote the function thereof, either as hardware, or as software, or as a combination thereof.

Fig. 15 shows an embodiment 200 similar to the embodiment 100, where instead of a separate SVM 3, that last SVM 3 is added into the Network Switch 2, thereby creating a new Network Switch 2'. A software computer program and/or hardware dedicated to perform the tasks of the SVM 3 are thus appended to the Network Switch 2 to implement the new Network Switch 2'. The hosts 1, or the hosts 1 via the user network 6' and the Internet 80. In the case of storage devices, the SVM Driver 60/400 will poll those storage devices 4 to retrieve the information through the storage network 5' and report of the status to the SVM 3/400 via the user network 6'and the Internet 80.

This embodiment 400 permits to emulate the SVM 3 in a remote location and to manage many SANs simultaneously. Evidently, the remote SVM 3/400 may be connected to any user network 6' such as an Ethernet network, or an Internet 80, or to any other network.

A further preferred embodiment 500, denominated as an Upgraded Network Switch (UNS 2*), is achieved by taking advantage of the processing capabilities inherent to Enhanced Network Switches 2E, like the Catalyst 5000/5500 Family Switches, made by CISCO and for which details are provided at the Internet address http//:www.cisco.com. The UNS 2*, is presented in Figs. 18 and 19. The configuration of the SAN in the embodiment 500 is similar of that of the embodiment 200 shown in Fig. 15. As before, the hosts 1 are linked by user network links 5 to the storage devices 4 via the UNS 2*, which is also connected by a user network link to the user network 6'.

The Upgraded Network Switch 2* (UNS 2*), seen in more detail in Fig. 19, incorporates a modified SVM 3, designated as the adjusted SVM 3/500. The adjusted SVM 3/500 may take advantage of the existing hardware components of the UNS 2* and be implemented therein, perhaps even by a limited addition of hardware and the necessary software.

Reference is made to Fig. 19. An Enhanced Network Switch 2E, such as the Catalyst 5000/5500, is shown to feature host-connection ports 1* and storage-device-connection ports 4* for coupling to, respectively, the hosts 1 and the storage devices 4. These connection ports 1* and 4* are coupled by storage network links 5 to the storage network 5'. The Catalyst 5000/5500 is manufactured with a processor 66, that may be replaced by a processing instance 66, coupled to each connection port, thus with one processing instance 66 for at least one host 1. The processing instance 66 may thus be implemented either as a processor running a computer program, or as dedicated hardware, to operate the SVM Driver 60. In view of the fact that the SVM Driver 60 imposes but a very light processing load, the operation of the SVM Driver 60 is easily relegated to the processing instance 66. The UNS 2* thus incorporates the SVM 3 Driver(s) 60 associated with the array of hosts 1. In addition, each processing instance 66 is coupled to the adjusted SVM 3/500 by a link 14 and the adjusted SVM 3/500 is also connected by a user network link 6 to the user network 6'. Each single processing instance 66 independently operates an SVM Driver 60, to route data from a host 1 to a storage device 4 via the UNS 2*. This process occurs in parallel and concurrently for a multiplicity of processors 66 of hosts 1 and of storage devices 4.

Table updating is carried out by the SVM Driver 60 operating in each processing instance 66, via the storage network 5'. The device polling is operated for the hosts 1 and for the storage devices 4 by the adjusted SVM 3/500, also via the storage network 5'.

CLAIMS

1. A method for storage virtualization in a Storage Area Network (SAN) comprising an array of hosts (1) coupled to an array of storage devices (4) via a Network Switch (2) operative for routing storage I/O operations between the array of hosts and the array of storage devices, the storage devices having a storage capacity, and the array of hosts, the array of storage devices and the Network Switch being coupled together in a storage network (5'), and the array of hosts being coupled to a plurality of user workstations (7) on a user network (6'), the method being characterized by comprising the steps of:

coupling a Storage Virtualization Manager (SVM) (3) by a storage network link (5) to the Network Switch on the storage network, and coupling the SVM by a user network link (6) to the user network, the SVM being configured for virtualization of the storage capacity and for managing metadata, and

operating the SVM for virtualization of the storage capacity and for managing metadata,

whereby virtualization of the storage capacity and managing metadata are decoupled from routing storage I/O operations.

2. The method according to Claim 1, characterized by further comprising the step of:

operating a virtualization computer program comprising a first configuration portion operating in the SVM (3) and a second translation portion (60) operating in association with at least one host of the array of hosts.

3. The method according to Claim 2, characterized by further comprising the step of:

operating the second translation portion of the virtualization computer program in each host of the array of hosts.

4. The method according to Claim 2, characterized by further comprising the step of:

operating the second translation portion of the virtualization computer program on a processor associated with at least one host of the array of hosts.

5. The method according to Claim 2, characterized by further comprising the step of:

operating the first configuration portion and the second translation portion of the virtualization computer program in interactive operative association.

6. The method according to the Claims 2 to 5, characterized by further comprising the step of:

operating the virtualization computer program for supporting real time configuration adaptation of the SAN in response to a configuration change occurring in the array of hosts.

7. The method according to the Claims 2 to 5, characterized by further comprising the step of:

operating the virtualization computer program for supporting real time configuration adaptation of the SAN in response to a configuration change occurring in the array of storage devices.

forming an Upgraded Network Switch (UNS) (2*) to upgrade the Enhanced Network Switch (2E), the UNS integrating an adjusted Storage Virtualization Manager (SVM) (3/500) configured for and operative for virtualization of the storage capacity and for managing metadata, the (UNS) comprising a first Enhanced Network Switch portion and a second adjusted SVM portion,

coupling the second adjusted SVM portion by a storage network link (5) to the first Enhanced Network Switch portion and by a user network link (6) to the user network, and

operating the second adjusted SVM portion for virtualization of the storage capacity and for managing metadata,

whereby virtualization of the storage capacity and managing of metadata are decoupled from routing storage I/O operations.

16. The method according to Claim 15, wherein the UNS (2*) is characterized by further comprising the steps of:

coupling a plurality of ports of the Enhanced Network Switch comprising coupling a first portion of ports (1*) to the hosts (1) and a second portion of ports (4*) to the storage devices (4), and coupling by storage network link (5) to each port of the first portion of ports and to each port of the second portion of ports, respectively, at least one host and at least one storage device, and

coupling a processing instance (66) to each port of at least the first portion of ports, and to the second adjusted SVM portion.

17. The method according to Claim 15, characterized by further comprising the steps of:

operating a virtualization computer program comprising a first configuration portion operative on the second adjusted SVM portion and a second translation portion operative on the processing instance coupled to each port of the first portion of ports.

18. The method according to Claim 17, characterized by further comprising the steps of:

operating independently the second translation portion of the virtualization computer program on the processing instance coupled to each port of the first portion of ports.

19. The method according to Claim 17 characterized by further comprising the step of:

running the first configuration portion and the second translation portion of the virtualization computer program in interactive operative association.

20. The method according to the Claims 17 to 19, characterized by further comprising the step of:

operating the virtualization computer program to support real time configuration adaptation of the SAN in response to a configuration change occurring in the array of hosts.

21. The method according to the Claims 17 to 19, characterized by further comprising the step of:

and to each port of the second portion of ports, respectively, at least one host and at least one storage device, characterized by further comprising the steps of:

coupling a processing instance (66) to each port of at least the first portion of ports, and

linking each processing instance to the adapted SVM (3/600) via the user network 6'.

29. The method according to Claim 27, characterized by further comprising the step of:

operating a virtualization computer program comprising a first configuration portion operated by the adapted SVM and a second translation portion operated by the processing instance coupled to each port of the first portion of ports.

30. The method according to Claim 29, characterized by further comprising the steps of:

operating independently the second translation portion of the virtualization computer program on the processing instance coupled to each port of the first portion of ports.

31. The method according to Claim 29, characterized by further comprising the steps of:

operating the first configuration portion and the second translation portion of the virtualization computer program in interactive operative association.

32. The method according to the Claims 29 to 31, characterized by further comprising the step of:

operating the first portion of the virtualization computer program to support real time configuration adaptation of the at least one SAN, in response to a configuration change occurring in the array of hosts.

33. The method according to the Claims 29 to 31, characterized by further comprising the step of:

operating interactively the second portion with the first portion of the virtualization computer program to support real time configuration adaptation of the at least one SAN, in response to a configuration change occurring in the array of storage devices.

34. The method according to the Claims 29 to 31, characterized by further comprising the step of:

operating computer program control functions comprised in the virtualization computer program for management of storage virtualization and for configuration management of both the array of hosts and the array of storage devices of the at least one SAN.

35. The method according to Claim 34, characterized by further comprising the step of:

enabling a System Administrator to manage the computer program control functions by operating a workstation coupled to the user network.

36. The method according to Claim 34, characterized by further comprising the step of:

45. The system according to the Claims 39 to 42, characterized by further comprising:
computer program control functions comprised in the virtualization computer program being operated for management of storage virtualization and for configuration management of both the array of hosts and the array of storage devices.

46. The system according to Claim 45, characterized by further comprising:
a System Administrator for managing the computer program control functions by operating a workstation coupled to the user network.

47. The system according to Claim 45, characterized by further comprising:
at least one user application computer program for managing the computer program control functions.

48. The system according to Claim 45, characterized by further comprising:
at least one storage application computer program for managing the computer program control functions.

49. The system according to Claim 38, characterized by further comprising:
a joint housing configured for incorporation therein of both the Network Switch and the SVM.

50. The system according to Claim 38, characterized by further comprising:
one chosen host of the array of hosts being configured for incorporation therein of the SVM, and the SVM being coupled in operative association to the Network Switch via the storage network.

51. The system according to Claim 39, characterized by further comprising:
a host coupled to the user network and being configured to operate the configuration computer program in operative association with the Network Switch and with the translation portion of the virtualization computer program operating on at least one processor coupled to the storage network.

52. A system for a Storage Area Network (SAN) comprising an array of hosts (1), an array of storage devices (4) having a storage capacity, and an Enhanced Network Switch (2E) operative for routing I/O operations between the array of hosts and the array of storage devices, the array of hosts being coupled to the array of storage devices via the Enhanced Network Switch, the array of storage devices and the Enhanced Network Switch being coupled together in a storage network (5'), and the array of hosts being coupled to a user network (6') comprising a plurality of user workstations (7), the system being characterized by comprising:

an Upgraded Network Switch (UNS) 2* created to upgrade the Enhanced Network Switch, the UNS integrating an adjusted Storage Virtualization Manager (SVM) (3/500) configured for and operative for virtualization of the storage capacity and for managing metadata, the UNS comprising a first Upgraded Network Switch portion and a second adjusted SVM portion, and

a storage network link (5) coupling the first Enhanced Network Switch portion to the second adjusted SVM portion and a user network link (6) coupling the second adjusted SVM portion to the user network,

whereby virtualization of the storage capacity and managing metadata are decoupled from routing storage I/O operations.

61. The system according to Claim 48, characterized by further comprising: at least one user application computer program for managing the computer program control functions.

62. The system according to Claim 48, characterized by further comprising: at least one storage application computer programs operating the management of the computer program control functions.

63. The system according to Claim 41, characterized by further comprising: a joint housing configured for incorporation therein of the first Enhanced Network Switch portion and the second adusted SVM portion.

64. A system for storage virtualization of at least one Storage Area Network (SAN) comprising an array of hosts (1), an array of storage devices (4) having a storage capacity, and a Enhanced Network Switch (2E) operative for routing storage I/O operations between the array of hosts and the array of storage devices, the array of hosts being coupled to the array of storage devices via the Enhanced Network Switch, with the array of hosts and the array of storage devices and the Enhanced Network Switch being coupled together in a storage network(5'), and the array of hosts being coupled to a user network (6') comprising a plurality of user workstations (7) and further linked to a remote host (1R) via an Internet (80), the system being characterized by comprising:

an adapted SVM (3/600) coupled in operative association with the remote host, the adapted SVM being configured for virtualization of the storage capacity and for managing metadata of the at least one SAN via the Internet and the user network, with the adapted SVM being operated for virtualization of the storage capacity and for managing metadata of the at least one SAN via the Internet and the user network, and

a coupling for linking the Enhanced Network Switch to the user network, whereby virtualization of the storage capacity and managing of metadata are decoupled from routing storage I/O operations.

65. The system according to Claim 64 wherein the Enhanced Network Switch (2M) comprises a plurality of ports having a first portion of ports (1*) for coupling to the hosts (1) and a second portion of ports (4*) for coupling to the storage devices (4), and coupling by storage network links (5) to each port of the first portion of ports and to each port of the second portion of ports, respectively, at least one host of the array of hosts and at least one storage device of the array of storage devices, the system being characterized by further comprising:

a processing instance (66) coupled to each port of at least the first portion of ports, and each processing instance being linked to each port out of the first portion of ports via the user network to the adapted SVM (3/600).

66. The system according to Claim 64, characterized by further comprising: a virtualization computer program comprising a first configuration portion operative on the adapted SVM and a second translation portion operative on the processing instance coupled to each port of the first portion of ports.

76. The storage virtualizer according to Claim 75, characterized by further comprising:
a virtualization computer program comprising a first configuration portion running on the SVM and a second translation portion operating in association with at least one host of the array of hosts.

77. The storage virtualizer according to Claim 76, characterized by further comprising:
each host of the array of hosts operatinh the second translation portion of the virtualization computer program.

78. The storage virtualizer according to Claim 76, characterized by further comprising:
a processor associated with at least one host of the array of hosts for operating the second translation portion of the virtualization computer program.

79. The storage virtualizer according to Claim 76, characterized by further comprising:
the first configuration portion and the second translation portion of the virtualization computer program being coupled in interactive operative association.

80. The storage virtualizer according to the Claims 76 to 79, characterized by further comprising:
a real time configuration adaptation of the SAN in response to a configuration change occurring in the array of hosts, the real time configuration adaptation being supported by operation of the virtualization computer program.

81. The storage virtualizer according to the Claims 76 to 79, characterized by further comprising:
a real time configuration adaptation of the SAN in response to a configuration change occurring in the array of storage devices, the real time configuration adaptation being supported by operation of the virtualization computer program.

82. The storage virtualizer according to the Claims 76 to 79, characterized by further comprising:
computer program control functions comprised in the virtualization computer program being operated for management of storage virtualization and for configuration management of both the array of hosts and the array of storage devices.

83. The storage virtualizer according to Claim 82, characterized by further comprising:
a System Administrator for managing the computer program control functions by operating a workstation coupled to the user network.

84. The storage virtualizer according to Claim 82, characterized by further comprising:
at least one user application computer program for managing the computer program control functions.

at least one processing instance (66) being coupled to each port of at least the first portion of ports, and to the second SVM portion.

91. The storage virtualizer according to Claim 89, characterized by further comprising:
a virtualization computer program comprising a first configuration portion operative on the second SVM portion and a second translation portion operative on the at least one processor coupled to each port of the first portion of ports.

92. The storage virtualizer according to Claim 89, characterized by further comprising:
each at least one processor coupled to each port of the first portion of ports independently operating the second translation portion of the virtualization computer program.

93. The system according to Claim 89, characterized by further comprising:
an interactive operative association coupling the first configuration portion and the second translation portion of the virtualization computer program.

94. The system according to the Claims 91 to 93, characterized by further comprising:
a real time configuration adaptation of the SAN in response to a configuration change occurring in the array of hosts, the real time configuration adaptation being supported by operation of the virtualization computer program.

95. The system according to the Claims 91 to 93, characterized by further comprising:
a real time configuration adaptation of the SAN in response to a configuration change occurring in the array of storage devices, the real time configuration adaptation being supported by operation of the virtualization computer program.

96. The system according to the Claims 91 to 93, characterized by further comprising:
computer program control functions comprised in the virtualization computer program for management of storage virtualization and for configuration management of both the array of hosts and the array of storage devices.

97. The system according to Claim 98, characterized by further comprising:
a System Administrator for managing the. computer program control functions .by operating a workstation coupled to the user network.

98. The system according to Claim 98, characterized by further comprising:
at least one user application computer program for managing the computer program control functions.

99. The system according to Claim 98, characterized by further comprising:
at least one storage application computer programs operating the management of the computer program control functions.

100. The system according to Claim 98, characterized by further comprising:
a joint housing configured for incorporation therein of the first Enhanced Network Switch portion and the second SVM portion.

101. A storage virtualizer for storage virtualization of at least one Storage Area Network (SAN) comprising an array of hosts (1), an array of storage devices (4) having a storage capacity, and a Enhanced Network Switch (2E) operative for routing

a real time configuration adaptation of the at least one SAN in response to a configuration change occurring in the array of storage devices, the real time configuration adaptation being supported by operation of the virtualization computer program.

108. The storage virtualizer according to the Claims 103 to 105, characterized by further comprising:
computer program control functions comprised in the virtualization computer program, for management of storage virtualization, and for management of both the array of hosts and the array of storage devices of the at least one SAN.

109. The storage virtualizer according to Claim 71, characterized by further comprising:
a System Administrator for managing the computer program control functions via a workstation (8) coupled to the user network.

110. The storage virtualizer according to Claim 71, characterized by further comprising:
the computer program control functions being managed by at least one user application computer program operating on a host of the at least one SAN.

111. The storage virtualizer according to Claim 71, characterized by further comprising:
the computer program control functions being managed by at least one storage application computer program operating on a host of the at least one SAN.

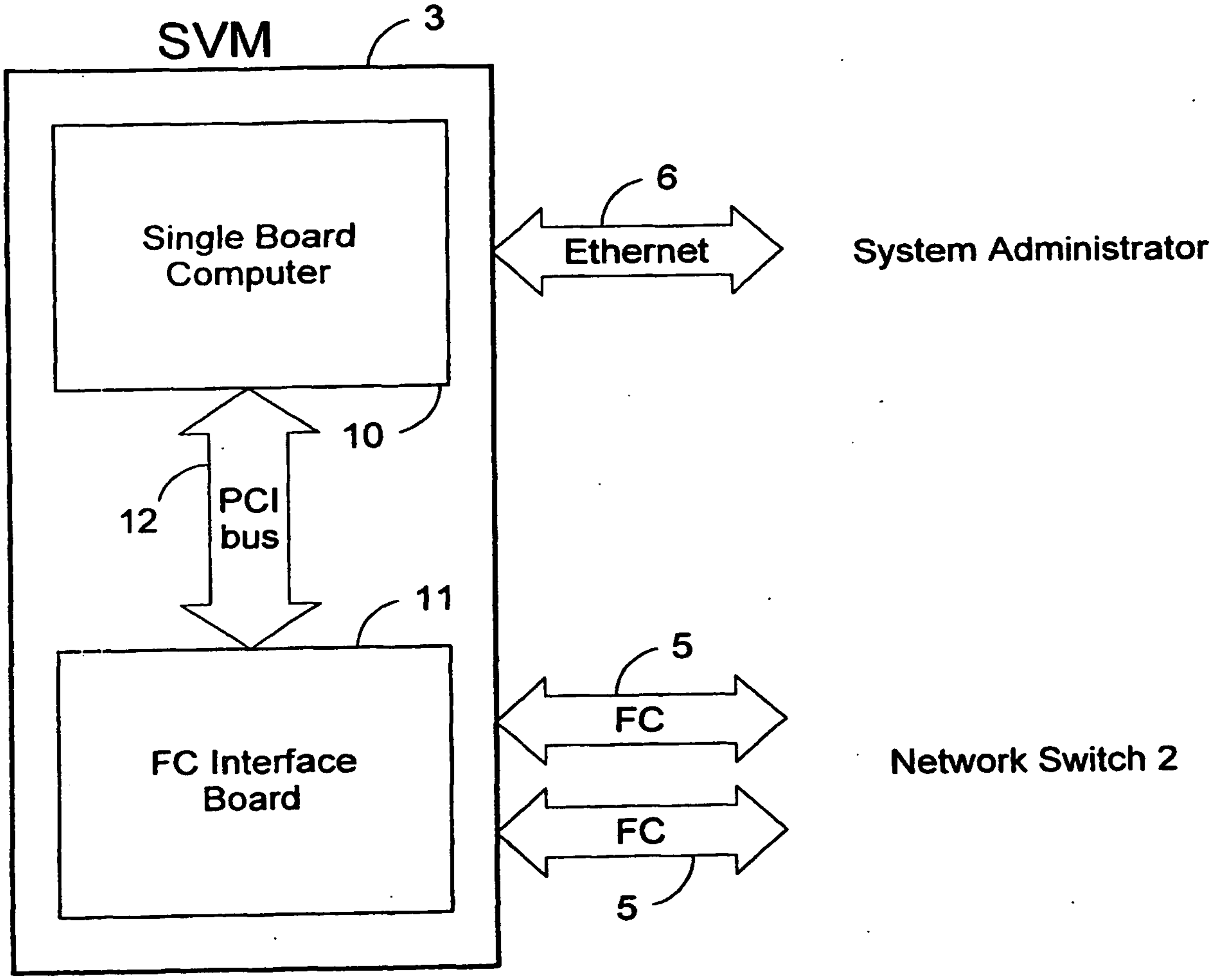

SVM
3
Single Board Computer
6
Ethernet
System Administrator
10
12
PCI bus
11
5
FC
Network Switch 2
FC Interface Board
FC
5

FIG.2

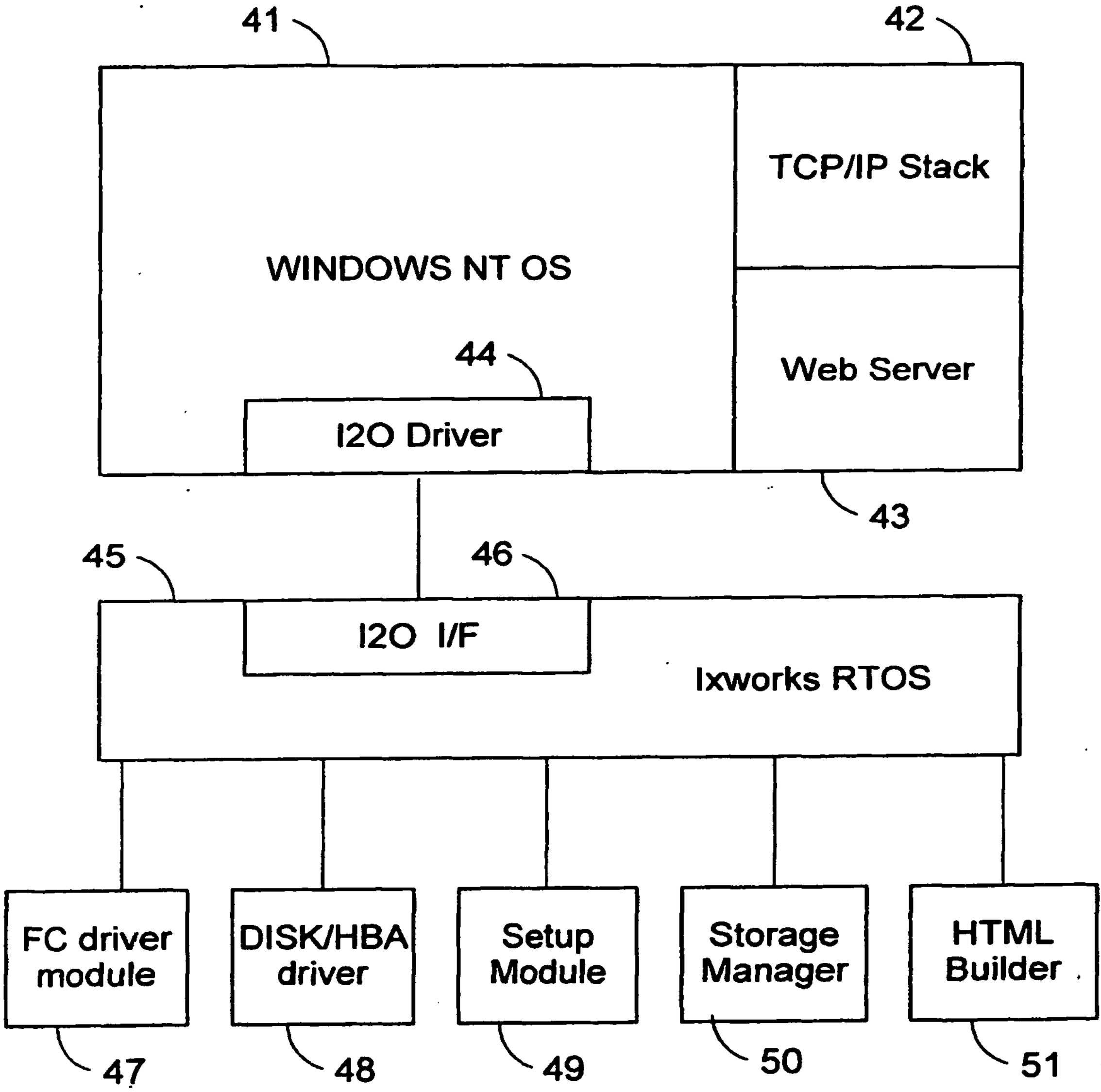

41
42
TCP/IP Stack
WINDOWS NT OS
Web Server
44
I2O Driver
43
45
46
I2O I/F
Ixworks RTOS
FC driver module
DISK/HBA driver
Setup Module
Storage Manager
HTML Builder
47
48
49
50
51

FIG.4

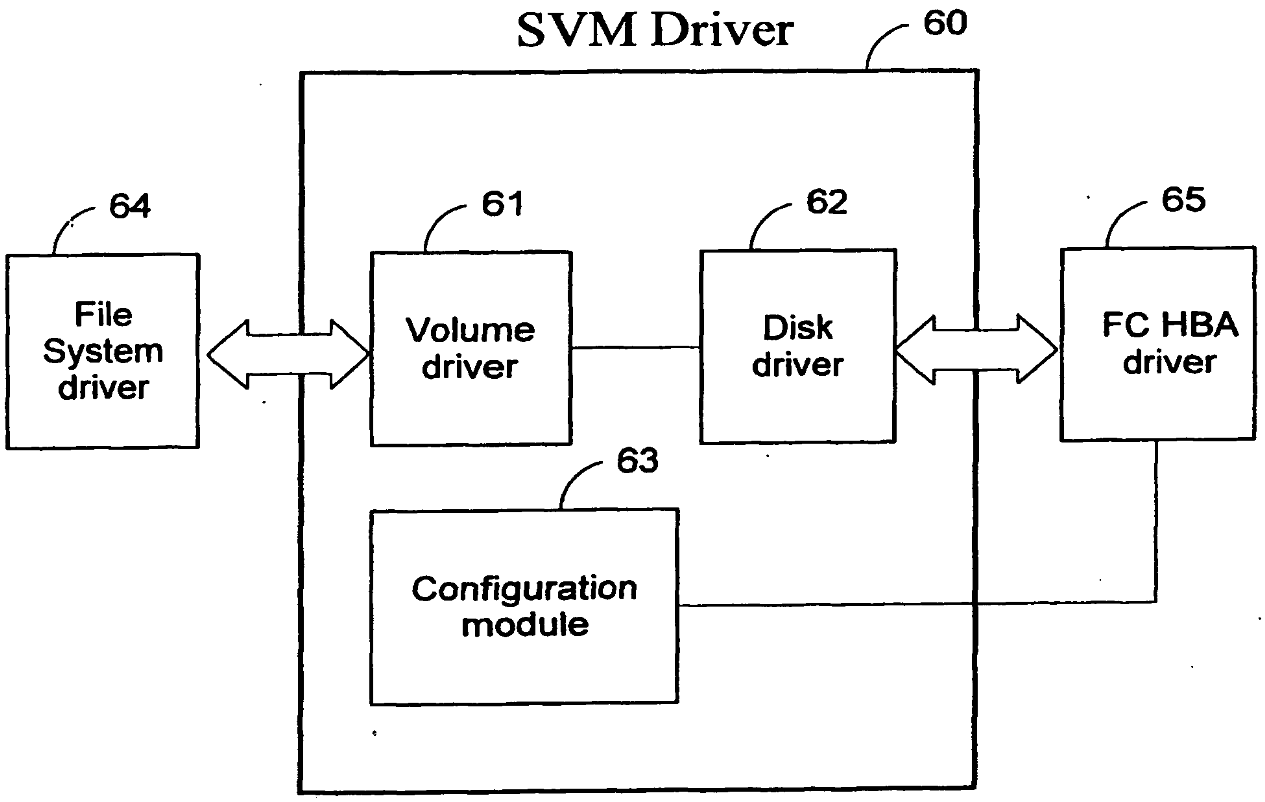

SVM Driver
60
64
File System driver
61
Volume driver
62
Disk driver
65
FC HBA driver
63
Configuration module

FIG.8

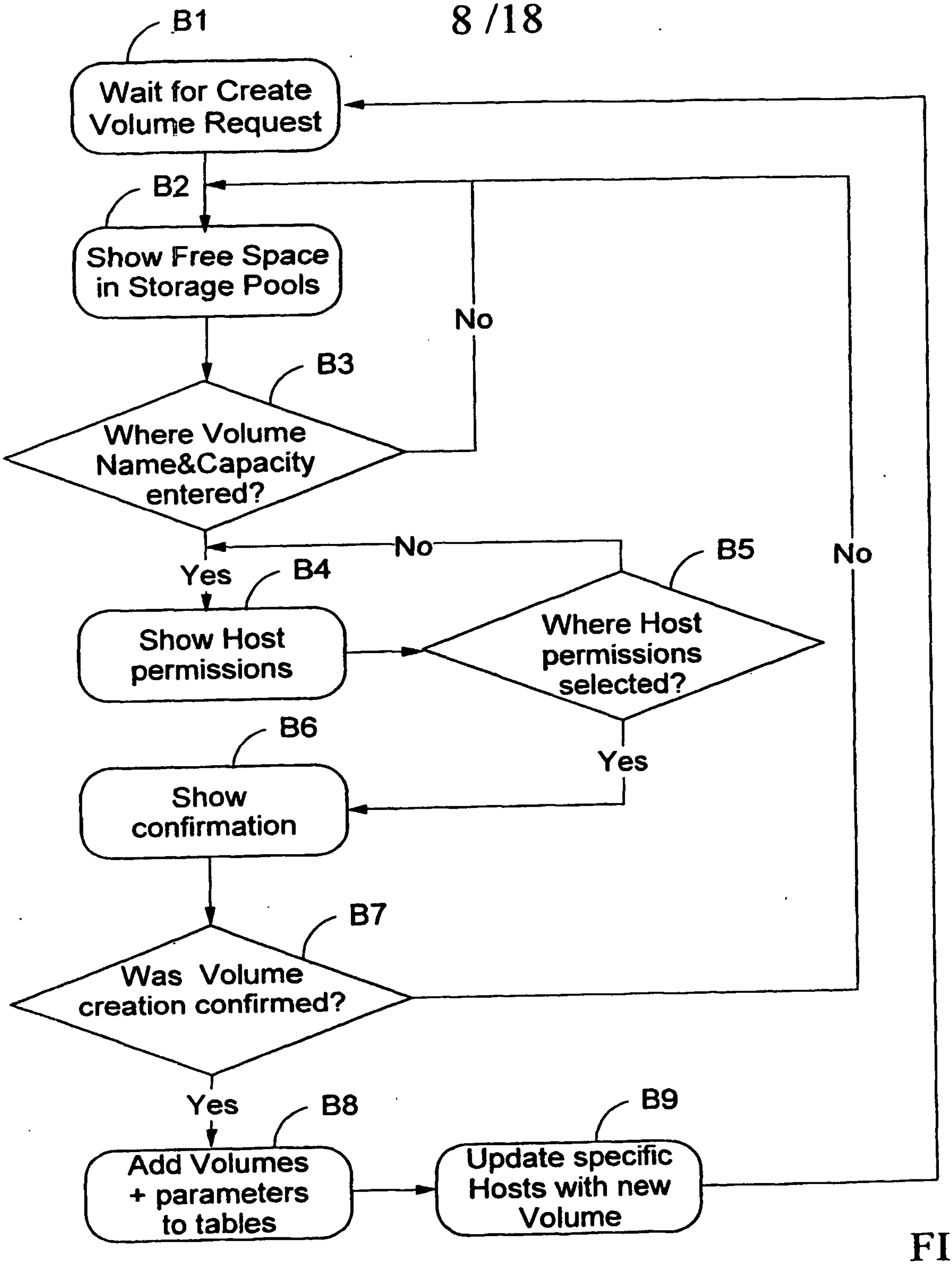

B1
Wait for Create Volume Request
B2
Show Free Space in Storage Pools
No
B3
Where Volume Name&Capacity entered?
Yes
No
B4
Show Host permissions
B5
Where Host permissions selected?
No
Yes
B6
Show confirmation
B7
Was Volume creation confirmed?
Yes
B8
Add Volumes + parameters to tables
B9
Update specific Hosts with new Volume

FIG.10

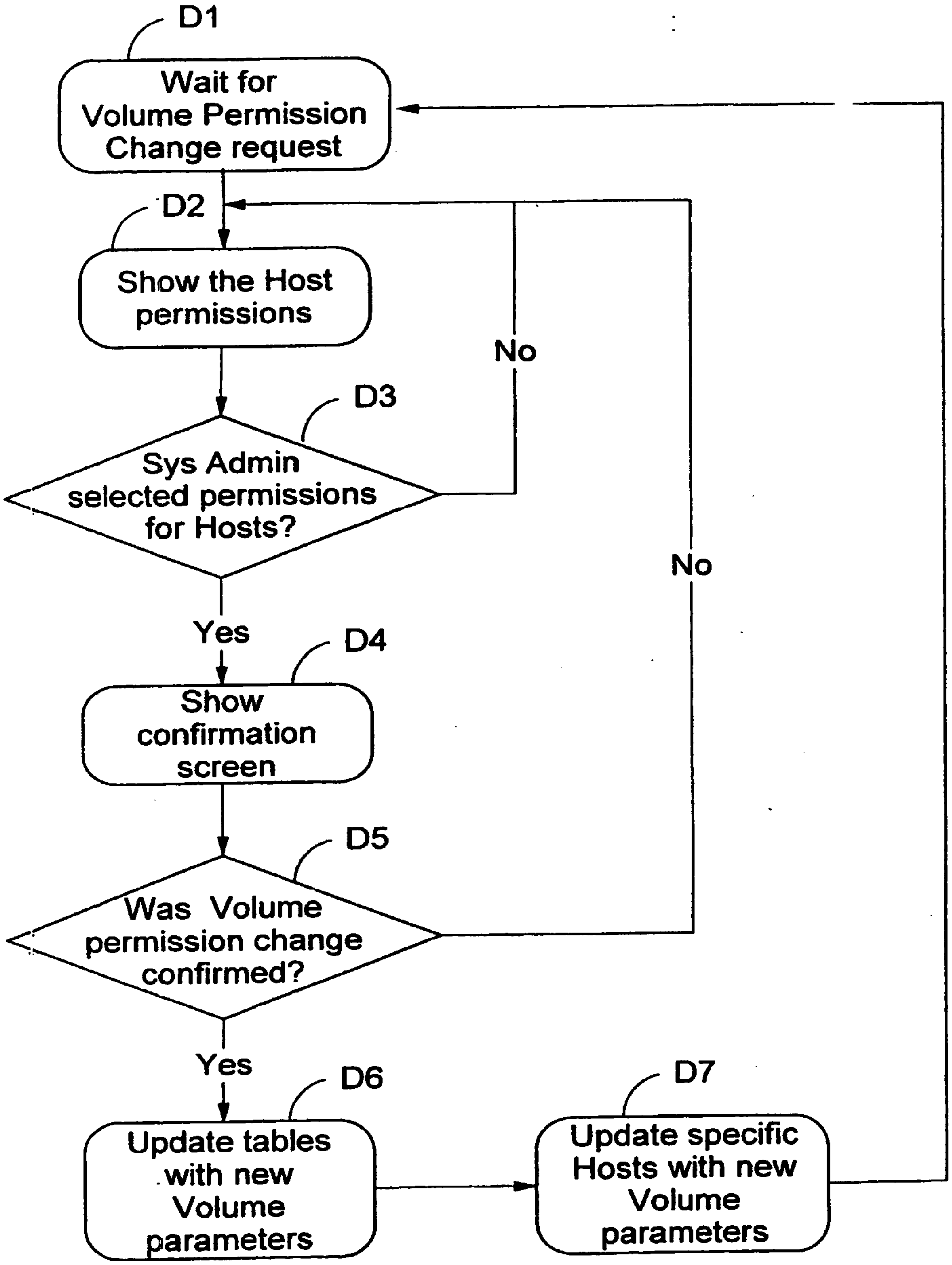

D1
Wait for Volume Permission Change request
D2
Show the Host permissions
No
D3
Sys Admin selected permissions for Hosts?
No
Yes
D4
Show confirmation screen
D5
Was Volume permission change confirmed?
Yes
D6
Update tables with new Volume parameters
D7
Update specific Hosts with new Volume parameters

FIG.12

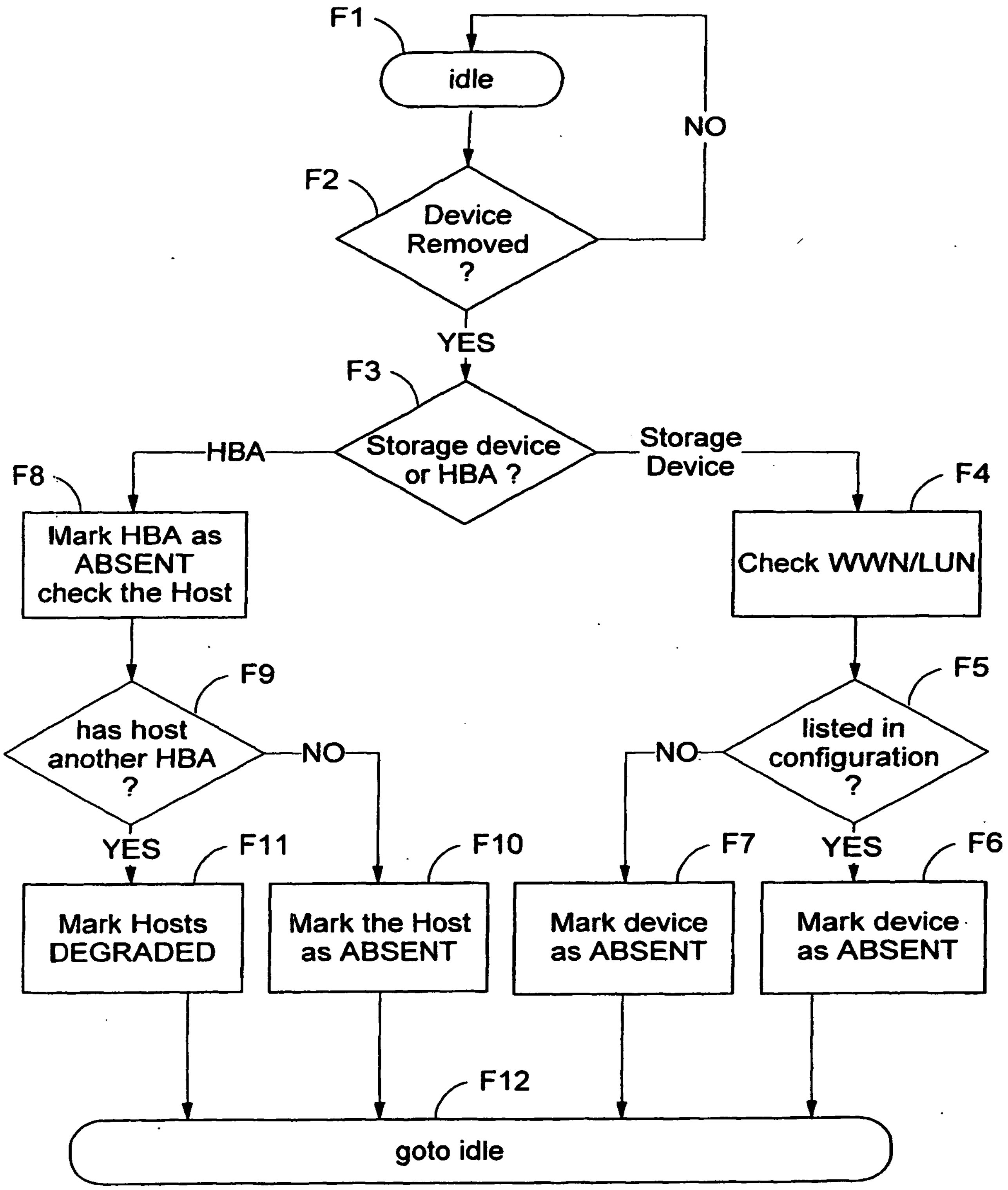

F1
idle
NO
F2
Device Removed ?
YES
F3
Storage device or HBA ?
HBA
Storage Device
F8
Mark HBA as ABSENT check the Host
F4
Check WWN/LUN
F9
has host another HBA ?
NO
F5
listed in configuration ?
NO
YES
F11
F10
F7
YES
F6
Mark Hosts DEGRADED
Mark the Host as ABSENT
Mark device as ABSENT
Mark device as ABSENT
F12
goto idle

FIG.14

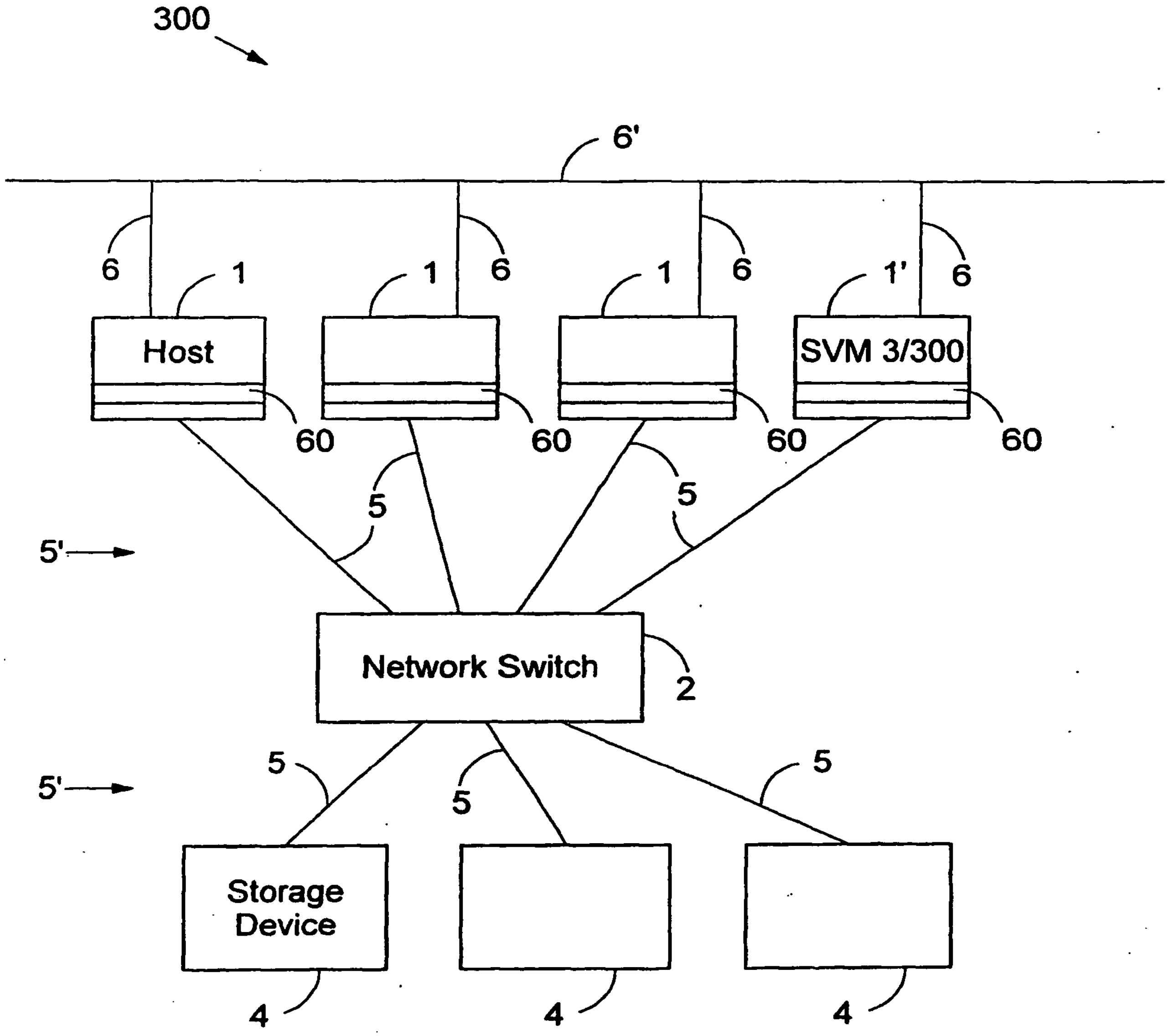

300
6'
6
1
1
6
1
6
1'
6
Host
SVM 3/300
60
60
60
60
5
5
5'
Network Switch
2
5
5'
5
5
Storage Device
4
4
4

FIG.16

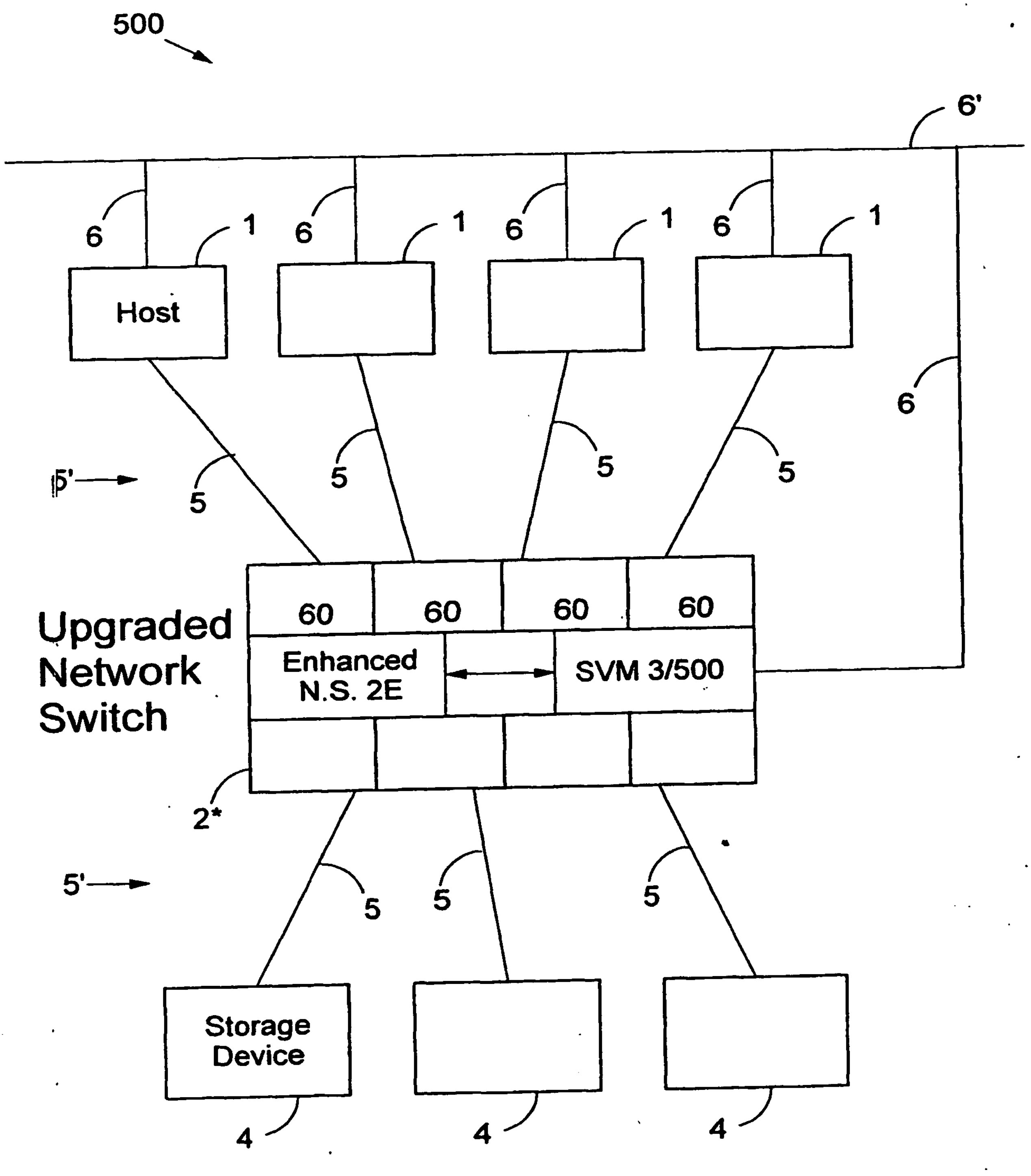

500
6'
6
1
6
1
6
1
6
1
Host
6
5'
5
5
5
5
Upgraded
Network
Switch
60
60
60
60
Enhanced
N.S. 2E
SVM 3/500
2*
5'
5
5
5
Storage
Device
4
4
4

FIG.18

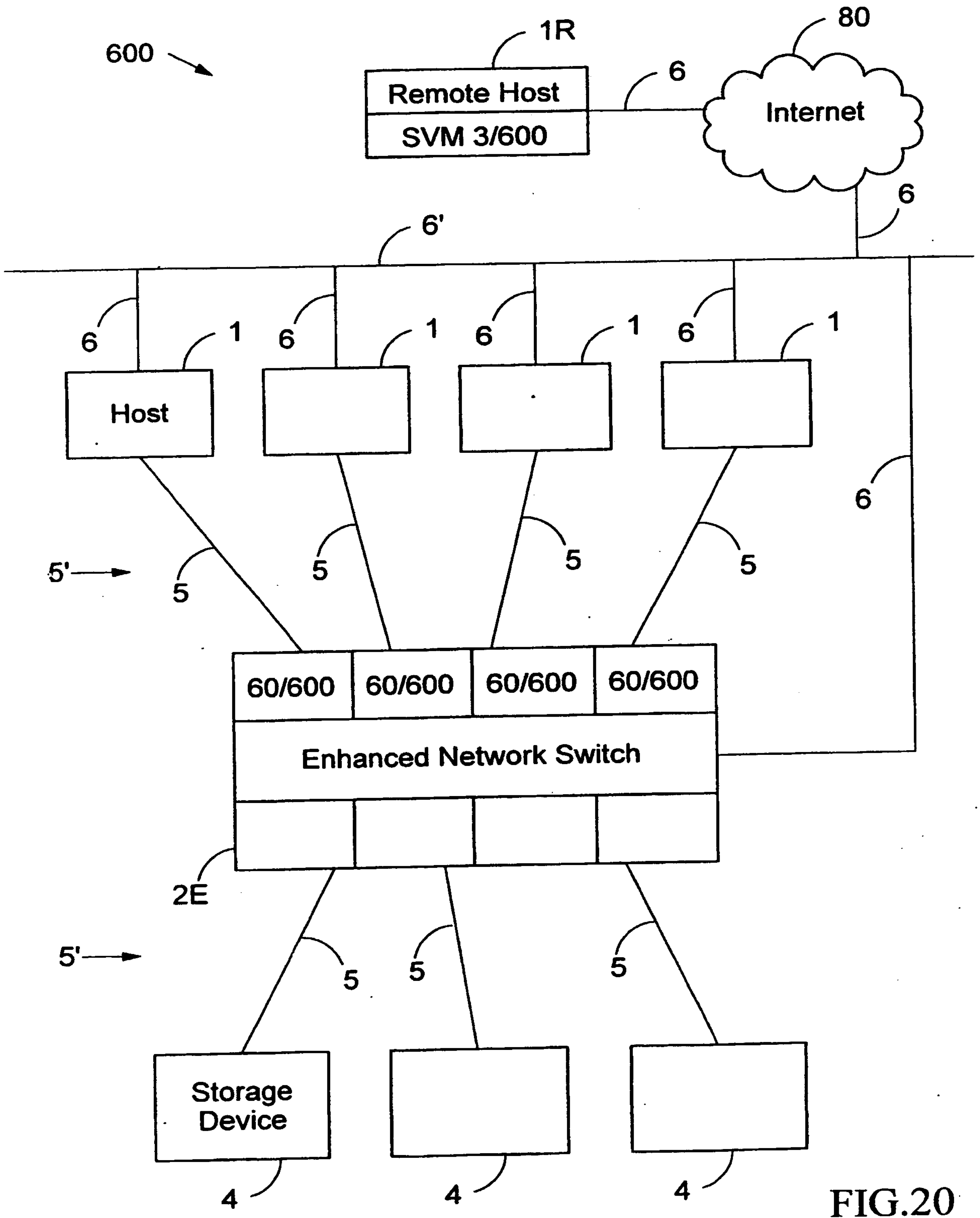

600
1R
Remote Host
SVM 3/600
6
80
Internet
6
6'
6
1
Host
6
1
6
1
6
1
6
5'
5
5
5
5
60/600
60/600
60/600
60/600
Enhanced Network Switch
2E
5'
5
5
5
Storage Device
4
4
4

FIG.20